

## Identify Entrepreneurial Opportunity by Mining Data from Social Media

Muhamad Shahbani Abu Bakar<sup>1\*</sup>, Izwan Nizal Mohd Shaharane<sup>2</sup>, Syahida Hassan<sup>3</sup>, Azman Ta'a<sup>4</sup>, Azman Yasin<sup>5</sup>,  
Ang Jin Sheng<sup>6</sup>, Safiah Suhaimi<sup>7</sup> and Khoo Yin Yin<sup>8</sup>

*School of Computing*<sup>1,3,4,5</sup>, *School of Quantitative Sciences*<sup>2,6</sup>, *School of Government*<sup>7</sup>,  
*Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia*  
*Faculty of Management and Economics*<sup>8</sup>,  
*Universiti Pendidikan Sultan Idris, 35900 Kuala Kubu Bharu, Perak, Malaysia*

### Abstract

*Social media plays an important role in marketing specifically in digital marketing. Many entrepreneur and businesses start to promote their brand and items on the social media platform such as Instagram. The main objective of this study is to crawl data from Instagram and perform some simple analysis to get some entrepreneurial ideas. A system named Graduate Entrepreneur Opportunity Recognition System (GEORsys) is proposed in this study. Hashtags for custom words or sectors like agriculture, services, food and beverages (F&B), fashion, information, communication and technology (ICT) and logistics are used to crawl data from Instagram. Then, some simple visualizations are shown to conclude what keywords or hashtags are popular in Instagram.*

**Keywords:** data crawling, Instagram, opportunity recognition, data mining, entrepreneurship

### INTRODUCTION

Instagram is one of the most prominent mobile photo-sharing platform at around the world (Oh, Lee, Kim, Park, & Suh, 2016). The estimation from Abbott, Donaghey, Hare, and Hopkins (2013) around four billion photos and 100 million users using Instagram in the year 2013. Miles (2013) reveals to his readers that how to promote business through Instagram. Thus, marketing business or searching business idea through Instagram is new revolution to business owners or entrepreneurs.

An entrepreneur is defined as people who can recognize opportunity and build an association to pursue that opportunity (Bygrave & Hofer, 1992). McGarry (1944) suggests that unemployment will turn to self-employment as one of the feasible alternatives. Thus, entrepreneurship can be considered as a major driver of economic growth by increasing the manpower contribution to output. Furthermore, entrepreneurship usually brings more employment and innovation (Omer, 2014). However, the main issue arises when the high educated graduates in Malaysia are best fit for white-collar jobs and they do not acquire the relevant vocational skills that allow them to engage in entrepreneurship activity. Besides, entrepreneur should be innovative. On the other hand, finding a suitable idea to venture a business is quite hard for students that do not have working experiences. Above all, students should utilize the power of social media to look for entrepreneur idea and opportunity.

### LITERATURE REVIEW

#### Entrepreneurship

Cunningham and Lischeron (1991) believe that the combination of French verb “entreprendre” and German word “unternehmen” which carries meaning of “undertake” derives the word entrepreneurship. Schumpeter (1975) views entrepreneurs as the agent of change while the market is in equilibrium due to entrepreneurs are innovators that creates new products, services, methods, and business models for the market and brings about change. Peredo and McLean (2006) mention that the word “entrepreneur” is

defined as people who begins or builds a commercial enterprise, especially financial risk is involved. On the other hand, Gartner (1989) define an entrepreneur as a person who recognize opportunity and takes necessary actions that relate with the creation of an institution to seek for that opportunity. Cole (1946) stated that entrepreneur must acquire distinctive personal qualities besides taking risk by buying at fix value and selling at uncertain value.

### Opportunity Recognition

The skill to recognize an opportunity and turn this opportunity into a applicable business that generates value for stakeholders is an important characteristics of an entrepreneur (Shane & Venkataraman, 2000). Ardichvili, Cardozo, and Ray (2003) point out that recognizing opportunity is the basic step of developing business process as this need to undergoes through various evaluation process. Baron (2006) describes opportunity recognition as a cognitive process where a person deduces that some opportunities and need gaps in the market or customers are identified by that person.

### Application of web content mining

The world wide web contains huge volume of information that anyone can accessible from any time and place (Herrouz, Khentout, & Djoudi, 2013). Therefore, entrepreneur can utilize world wide web as a supply of information that they needed. However, data from the world wide web usually comes with many formats including semi-structured and unstructured format. These formats hard to extract directly and gain information (Herrouz et al., 2013). However, Upadhyay and Dhingra (2013) point out that web mining able to extract information from the enormous amount of data available on the web effectively and efficiently. Web mining can be divided into few categories including web content mining, web structure mining and web usage mining. One of the categories is web content mining. The main function of web content mining is to extract useful information from the web page content and find out any useful insights or patterns (Herrouz et al., 2013). (Herrouz et al., 2013). Li, Lei, Khadiwala, and Chang (2012) propose a system named TEDAS (Twitter Based Event Detection and Analysis System) to mine data from social media platform specifically Twitter. Neri, Aliprandi, Capeci, Cuadros, and By (2012) used Sentiment and Knowledge Mining system developed by Baldini, Neri, and Pettoni (2007) to perform sentiment analysis on Facebook. Abdul-Mageed, Diab, and Kübler (2014) present a system for Subjectivity and Sentiment Analysis (SSA) called SAMAR to mine Arabic social media's data. Salloum, Al-Emran, Monem, and Shaalan (2017) do a comprehensive study on text mining in social media specifically Facebook and Twitter. Younis (2015) mine Twitter data using an open source tools called R statistical software by performing text mining and sentiment analysis on the data crawled by R.

## METHODOLOGY

### GEORsys

Graduate Entrepreneur Opportunity System GEORsys is fully developed using R programming language. Shiny library is used to make the web application for GEORsys. The data crawling engine is adapted from Schröder (2018). Figure 1 shows the user interface of GEORsys.



Figure 1. User Interface of GEORsys

## Procedure Visualizing the Data

The procedures of analyzing data in this study which crawled by GEORsys are stated as below:

1. Firstly, choose all sector option in GEORsys and press the button of crawl the data.
2. Each keyword of the sectors in GEORsys including “agriculture”, “services”, “foodandbeverages”, “fashion”, “ICT”, “logistics” are crawled from Instagram for 10 seconds respectively.
3. Then, the data crawled from Instagram is combined.
4. The highest 20 frequency of hashtags found in the crawled data is presented using a bar graph.
5. Lastly, a word cloud is presented.

## RESULTS AND DISCUSSION

Figure 2 and Figure 3 illustrate the results of data crawled from GEORsys. Figure 2 is the word cloud and Figure 3 is the bar chart for the most frequent hashtags crawled using GEORsys. It can be concluded that logistics, services and fashion is the top 3 highest number of hashtags among the sectors crawled from Instagram.



Figure 2. Word Cloud for Most Frequent Hashtags

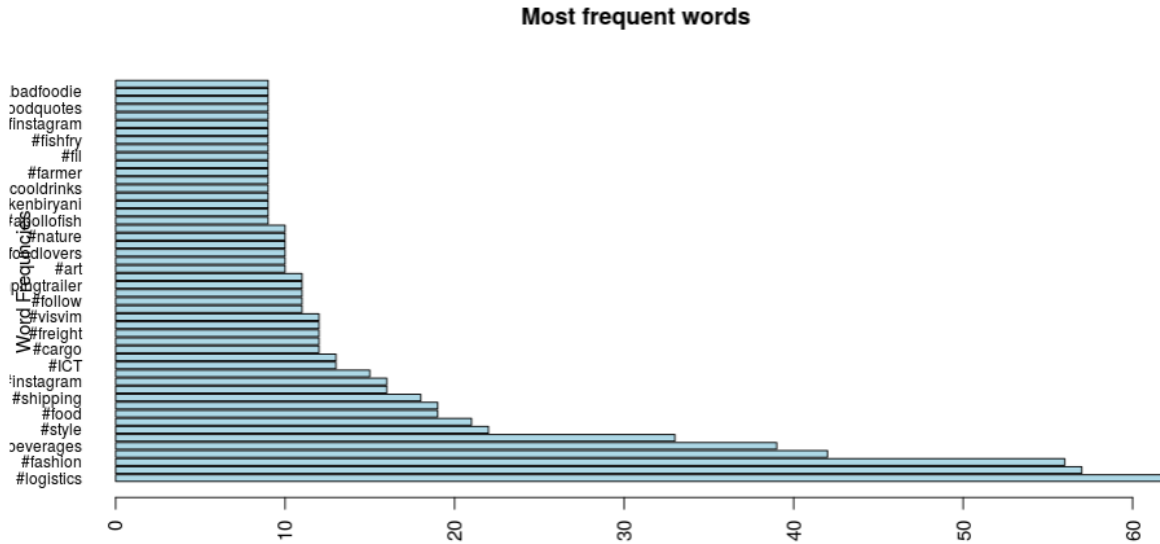


Figure 3. Bar Chart for Most Frequent Hashtags

## CONCLUSION

In brief, opportunity recognition is an essential process when begin a business. Lacking opportunity caused entrepreneurs hard to start one business. The aftermath of an incompetent opportunity recognition ability in venturing into a non-lucrative business will emerge into a premature closure. The means to promote opportunity recognition among student entrepreneurs, government should sponsor and promote more research in this area. Mining data from social media has been applied for a variety of purposes. However, purpose of web content mining to identify opportunity for entrepreneurs is not one of the extensively researched areas. GEORsys is proposed to help students and entrepreneurs who are lacking knowledge and experience in opportunity recognition. Nevertheless, the main limitation of this system is it can only produce initial business ideas for entrepreneurs but not providing the entrepreneurs the knowledge to evaluate and exploit the opportunity.

## ACKNOWLEDGMENTS

This research was supported by Ministry of Education (MOE) through Fundamental Research Grant Scheme (FRGS/1/2017/ICT04/UUM/02/5). We also want to thank to the Government of Malaysia which provide MyBrain12 program for sponsoring this work under the self-fund research grant and L00022 from Ministry of Science, Technology and Innovation (MOSTI).

## REFERENCES

- [1] Abbott, W., Donaghey, J., Hare, J., & Hopkins, P. (2013). An Instagram is worth a thousand words: An industry panel and audience Q&A. *Library Hi Tech News*.
- [2] Abdul-Mageed, M., Diab, M., & Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language*, 28(1), 20-37.
- [3] Ardichvili, A., Cardozo, R., & Ray, S. (2003). A theory of entrepreneurial opportunity identification and development. *Journal of Business venturing*, 18(1), 105-123.
- [4] Baldini, N., Neri, F., & Pettoni, M. (2007). A multilanguage platform for open source intelligence. *WIT Transactions on Information and Communication Technologies*, 38, 325-334.

- [5] Baron, R. A. (2006). Opportunity recognition as pattern recognition: How entrepreneurs “connect the dots” to identify new business opportunities. *Academy of management perspectives*, 20(1), 104-119.
- [6] Bygrave, W. D., & Hofer, C. W. (1992). Theorizing about entrepreneurship. *Entrepreneurship theory and Practice*, 16(2), 13-22.
- [7] Cole, A. H. (1946). An approach to the study of entrepreneurship: A tribute to Edwin F. Gay. *The Journal of Economic History*, 6(S1), 1-15.
- [8] Cunningham, J. B., & Lischeron, J. (1991). Defining entrepreneurship. *Journal of small business management*, 29(1), 45-61.
- [9] Gartner, W. B. (1989). Some suggestions for research on entrepreneurial traits and characteristics. *Entrepreneurship theory and Practice*, 14(1), 27-38.
- [10] Herrouz, A., Khentout, C., & Djoudi, M. (2013). Overview of web content mining tools. *arXiv preprint arXiv:1307.1024*.
- [11] Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C.-C. (2012). *Tedas: A twitter-based event detection and analysis system*. Paper presented at the 2012 IEEE 28th International Conference on Data Engineering.
- [12] McGarry, E. (1944). New Firms and Free Enterprise. In: JSTOR.
- [13] Miles, J. (2013). *Instagram power*: McGraw-Hill Publishing.
- [14] Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012). *Sentiment analysis on social media*. Paper presented at the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
- [15] Oh, C., Lee, T., Kim, Y., Park, S., & Suh, B. (2016). *Understanding participatory hashtag practices on Instagram: a case study of weekend hashtag project*. Paper presented at the Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.
- [16] Omer, M. M. (2014). Significance of Entrepreneurship? A Appraisal of Recent Research. *Significance*, 6(14).
- [17] Peredo, A. M., & McLean, M. (2006). Social entrepreneurship: A critical review of the concept. *Journal of world business*, 41(1), 56-65.
- [18] Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: facebook and twitter perspectives. *Adv. Sci. Technol. Eng. Syst. J*, 2(1), 127-133.
- [19] Schröder, J. (2018). Build Your Own Instagram Database. Retrieved from <https://towardsdatascience.com/build-your-own-instagram-database-134281e8ee92>
- [20] Schumpeter, J. A. (1975). Capitalism, socialism and democracy (1942). *J. Econ. Literature*, 20, 1463.
- [21] Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of management review*, 25(1), 217-226.
- [22] Upadhyay, G. M., & Dhingra, K. (2013). Web Content Mining: Its Techniques and Uses. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(11), 610-613.
- [23] Younis, E. M. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study. *International Journal of Computer Applications*, 112(5).